DOCUMENT RESUME

ED 371 034 TM 021 738

AUTHOR Gooding, Kathleen
TITLE Teaching to the Test: The Influence of Alternative Modes of Assessment on Teachers' Instructional Strategies.
PUB DATE Apr 94
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Research; Elementary Education; *Elementary School Teachers; Knowledge Level; Questionnaires; *Research Utilization; Surveys; *Teacher Education; *Teaching Methods; Training
IDENTIFIERS *Alternative Assessment; Reform Efforts; Self Regulation; Self Report Measures; *Teaching to the Test

ABSTRACT

A survey of teachers in New York provides an initial step toward the development of a framework for defining, with some consistency, commonalities among the various methods of assessment, as it identifies the instructional strategies self-reported by teachers trained in a variety of alternative assessment methods. Elementary school teachers (n=191) responded to a questionnaire developed for the study and validated through expert evaluations. Teacher responses were analyzed according to the level of training in alternative assessment and the perceived level of implementation. Significant differences were found in the instructional practices of teachers implementing and those not implementing alternative modes of instruction. Self-reported strategies of implementing teachers seem to incorporate the use of research-based behaviors and practices believed to affect student learning, application of knowledge, and self-regulated behavior positively. Assessment appears congruent with instruction. In addition, training and perceived support seem important to teachers' use of strategies. Nine tables and seven figures present study findings. (Contains 54 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 371 034

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

KATHLEEN GOODING

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Title: Teaching to the Test: The Influence of Alternative Modes of Assessment on Teachers' Instructional Strategies

Author: Kathleen Gooding
4 Pennington Drive
Huntington, New York 11743
(516)673-9299

Institution: New York University, NYC, N.Y.

1994 American Educational Research Association Annual Meeting
New Orleans, Louisiana
April 4, 1994

BEST COPY AVAILABLE

TM021738

TEACHING TO THE TEST: THE INFLUENCE OF ALTERNATIVE MODES OF ASSESSMENT ON TEACHERS' INSTRUCTIONAL STRATEGIES

Introduction

Background of the Problem

> It is axiomatic that measurement drives instruction: What is tested is what is taught. In addition, the way a subject is tested influences the way in which it is taught. Consequently, student assessment has long been the focal point of educational reform, and it will continue to be so during the 1990s. (Lott, 1991)

Accountability and increased documentation of what students know and are able to do are among the issues discussed on educational reform agendas (Gardner, 1989; Neill & Medina, 1989; Resnick & Resnick, 1989a; Simmons & Resnick, 1993). A debate exits between advocates of traditional, standardized or norm referenced methods of assessment and the advocates of alternative modes of assessment. The debate focuses on the ability of the various means of assessment to provide adequate and valid measures of student accomplishment. (Aschbacher, 1991; National Commission on Testing and Public Policy, 1990). Advocates of alternative modes of assessment argue that evaluative measures should have systemic validity (Frederiksen & Collins, 1989). That is, they should teach as well as test. Assessment should reflect the best educational practices, reflect real-life challenges of the domain, be responsive to different learning styles and intelligences, measure student mastery of the subject, document personal growth over time and be criterion or curriculum referenced, rather than norm referenced (Gardner, 1989; Herman, Aschbacher, & Winters, 1992; Wolfe, Bixby, Glenn, & Gardner, 1991). If alternative modes of assessment can live

up to these expectations, it could have tremendous influence on classroom practice and environment. Nevertheless, training teachers to use new instructional or assessment strategies, or mandating their use, does not mean that those strategies will become part of classroom practice (Lieberman & Miller, 1991).

As the debate continues, many districts and states have moved toward implementing alternative assessment measures (Aschbacher, 1991). This study begins to look at the impact this present reform movement in assessment and evaluation is having on the classroom situation. Using survey methodology, the study compares the instructional strategies and practices reported used in the classroom by four groups of teachers with various levels of training in and self reported use of alternative modes of assessment.

Review of the Problem

This study provides an initial step toward the development of a framework for defining, with some consistency, the commonalities between the various methods of assessment and identifies the instructional strategies self-reported by teachers trained in a variety of the alternative assessment methods within New York State. It then attempts to identify those factors which may contribute to the implementation of those instructional strategies most closely related to the components of alternative modes of assessment.

Before the survey could be initiated, it was necessary to formulate a common definition for alternative modes of assessment. To do this the basic guidelines or

dimensions of the most recognized alternative assessment models were transferred into a common set of criteria which would accurately represent the theoretical framework of the underlying philosophy without corrupting the individual focus of specific methods (Baker & Stites, 1991; Baron, 1990a; Stiggins, 1989;Wiggins, 1991b; Wolf, 1989). Through careful analysis of the major alternative assessment methodologies, seven commonalities or components were identified: longitudinal, authentic, non-secretive, non-biased, integrated knowledge, process-oriented, and student-focused (Table 1). These components were operationalized to form the basis of the questionnaire developed for the survey.

Classification of teachers' responses by descriptors of the common definition enabled analysis of the instructional strategies and identification of those strategies most closely related to implementation of alternative assessment practices in the classroom. By identifying and clarifying the instructional strategies perceived to be used by teachers in the field, behavioral trends were identified and isolated. Using these isolated trends, teachers were categorized as implementors and non-implementors of alternative assessment practices. Comparisons were made to determine how training and perceived implementation of alternative modes of assessment displayed themselves in the instructional strategies reported by teachers.

Theoretical or Conceptual Framework

A search of the literature revealed a strong theoretical basis supporting the shift in assessment methodologies from standardized, norm-referenced testing to more

Table 1

Common Components of Alternative Modes Of Assessment

COMPONENT	DESCRIPTION
Longitudinal	-measures student achievements over a given period of time -multi-staged, leading to refined and revised products or performances
Authentic	-reflects tasks and thinking processes that simulate, or parallel the kinds of challenges necessary to succeed in that field of study outside of the school environment as an adult
Not Secretive	-based on **standards**, known in advance, to which the student strives -promotes interaction between students and adults, use of many resources, editing, and improvement before final evaluation
Non-biased	-appropriate to different learning styles, interests, backgrounds, and cultures -addresses specific needs and criteria of district, school, and individual
Integrates Knowledge	-involves the production, not reproduction of knowledge -requires use of a repertoire of knowledge, and application of more than one form of intelligence to the completion of the task
Process Oriented	-requires demonstration and documentation of knowledge growth, creativity, problem-solving, skill acquisition -focuses on the students' ability to produce a quality **PRODUCT OR PERFORMANCE**
Student Focused	-actively involves student in self assessment, peer assessment -actively engages student in process, development, evaluation and implementation of assessment

authentic, performance-based measures. The research offered extensive documentation of the effects of standardized, short answer or multiple-choice testing on the instructional strategies of teachers (Herman, Dreyfus & Golan, 1990; Herman & Golan, 1991; Neill & Medina, 1989; Shepard 1990). It offered limited empirical insight into the influence of the newer alternative methods of assessment on instructional strategies. Evidence that was available related to major training initiatives. Training initiatives that were strongly supportive both fiscally and philosophically by partnerships of schools and training organizations. Although results from some of these initiatives, like Arts Propel (Wolf, LeMahieu & Eresh, 1992) in the Pittsburgh Schools, were promising, others, like the Vermont Portfolio Project (Koretz, Stecher & Deibert, 1992), and Kentucky Instructional Results Information System (Vitali, 1993) gave first insights into the problems that face implementation of the methodology on a large scale basis.

Present efforts within New York State as related to the implementation of the New Compact for Learning (NYSED, 1990) require that each district take a close look at present instructional activities in the classroom, present assessment methods used and move toward a more performance-based, authentic system of assessment. Meanwhile, fiscal resources for most school districts in New York State continue to be limited if not restrictive. Although a few districts in the State have been fortunate enough to bring recognized consultants in alternative modes of assessment into their school environment for intensive work, most have not. Training in new assessment methods often results in sending teachers to a series of workshops or in-service

courses rather than infusion into the school environment and community. To plan for the adoption of these new assessment measures educators need to know how teachers internalize these differing perspectives of alternative modes of assessment and how they perceive themselves as adapting it to behaviors in the classroom situation.

Research Design

Three hundred and twenty teachers (K-8) in New York State were sampled. One hundred and ninety one usable questionnaires were returned for a response rate in excess of 60%. To identify the sample, consultants or organizations offering training in alternative modes of assessment identified districts developing alternative assessments. To create a matched sample minimizing the variation between trained and untrained teachers, equal numbers of trained and untrained teachers were surveyed within each participating district. When this was not possible, teachers in neighboring districts with similar demographic characteristics were surveyed. The final sample reflected equal distributions between the two groups of teachers on dimensions of grade taught, years teaching, class size, school organization, structure of assignment and size of school.

The Instrument

The identified components of alternative modes of assessment were operationalized into a specific series of student and teacher behaviors representative of instructional strategies used by teachers in the classroom. Questions based on the

operationalized behaviors formed the basis of the questionnaire used for the survey. Validation of the questionnaire was obtained through a series of expert evaluations and pilots. Reliability ratings for each of the component indexes are contained in Table 2.

Defining Training and Implementation

Table 2

Reliability of Alternative Assessment Components

Component	No. of items	Cronbach alpha
Student focused	15	.8796
Longitudinal	10	.7898
Integrates knowledge	10	.8406
Non-Biased	13	.7916
Process-oriented	14	.7848
Authentic	15	.8356
Not Secretive	13	.7984

Teacher responses were analyzed according to both the level of training in alternative modes of assessment and the perceived level of implementation. Teachers trained in alternative modes of assessment were initially identified by consultants and district administrators. This information was verified by teachers through a series of items on the questionnaire asking teachers to respond to specific information about

training in alternative modes of assessment. Specific items in the questionnaire believed to be closely related to implementation of alternative modes of assessment in the classroom were separated to form an implementation index. The implementing group was divided into trained and not trained subgroups. This enabled the identification of the four groups of teachers studied: Trained Implementors, Non-trained Implementors, Trained Non-implementors and Non-trained Non-implementors (Figure 1).

		LEVEL OF IMPLEMENTATION N =191	
		High Use of Alternative Assessment	Low/No Use of Alternative Assessment
		Implementing n = 83 % N = 44	Non-Implementing n =108 % N = 57
LEVEL OF TRAINING IN ALTERNATIVE MODES OF ASSESSMENT N =191	Training n =109 % N = 57	Trained Implementors n =57 % N =30	Trained Non-Implementors n =52 % N =27
	No Training n = 82 % N = 43	Non Trained Implementors n =26 % N =14	Non Trained Non-Implementors n =56 % N =29

Figure 1: Group Titles and n for Training and Implementation Matrix

The Sample

Table 3

Summary of Questionnaire Responses and District Demographic Statistics (N=320)

Questionnaire Distribution and Return					
# Districts	# Sent	# Return	% Return	#Trained	# Not Trained
13	320	194	61%	109	82

District Demographic Statistics				
	District Type	Pupil/ Teacher Ratio	% Minority	CAADM (per pupil exp.)
Range of Statistics	3-4	10-14	0-8	5130-12446
Total statistics		164	33	121346
Average per participating district	4	12.6	2.5	9334
State Average		13.8	13	6908
Suburban Average	4	13.4	3	7531

Of the 191 responses to the survey, 57% reported training in alternative modes of assessment. There were no statistically significant differences between teachers with and without training in alternative modes of assessment in regards to years

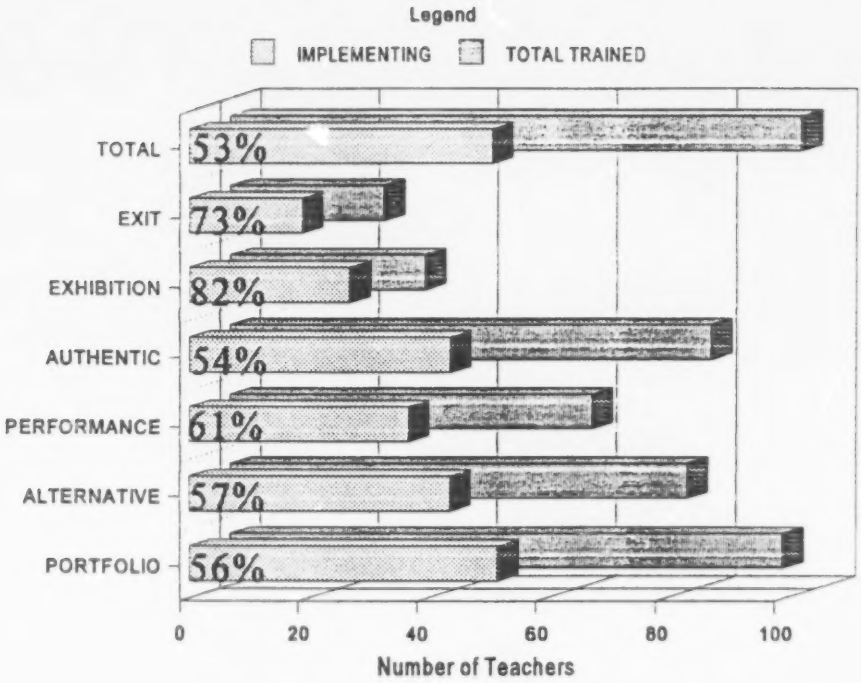

Figure 2. Number and Percent of Teachers Implementing According to Methods of Training in Alternative Modes of Assessment

find statistically significant differences in the amount of training for the two groups. Teachers implementing alternative assessment had more hours of training in every method (Figure 3).

Training in Innovative Instructional Methods

Since it was believed that training in other innovative methods of instruction could influence the reported use of strategies related to the perception of implementing alternative modes of assessment, the amount of training in five of the most popular innovative instructional methods were also analyzed. The innovative instructional

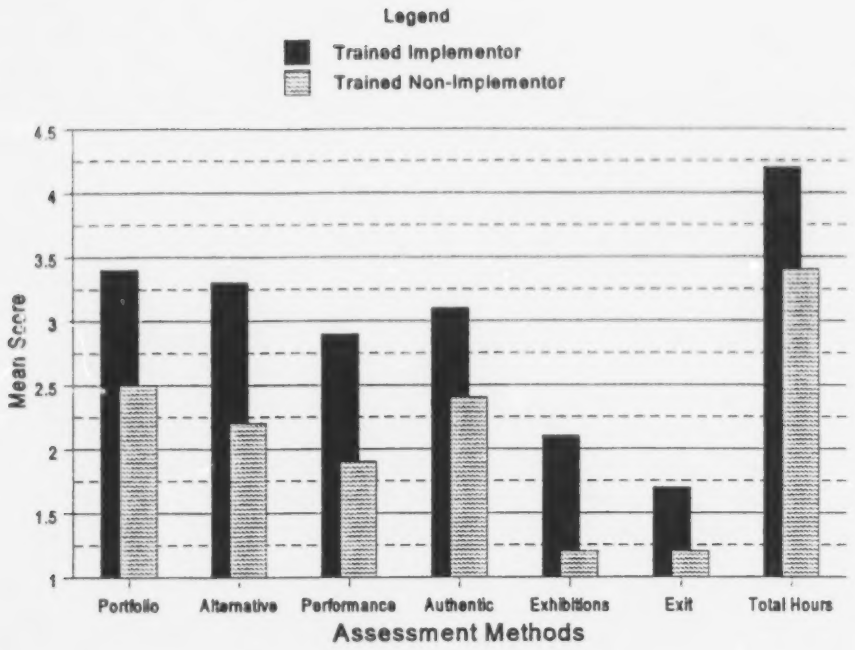

Figure 3. T-Test Comparison of Mean Hours of Training in Alternative Assessment Methods for Implementors and Non-Implementors

methods surveyed included: interdisciplinary instruction, whole language, literature-based thematic units, writing across the curriculum, "hands-on" science and cooperative learning. The majority (83%) of the teachers responding to the survey reported some training in one or more of the innovative methods. When the groups were categorized as implementing and non-implementing we see statistically significant differences between the Non-Trained Non-Implementors and the Trained Implementors, but not between the Non-Trained Implementors and the Trained Non-Implementors (Table 4)

Table 4

ANOVA Comparisons of Mean Scores for Training in Alternative Assessment and Innovative Instructional Methods

(N=191)

	Implementors n=83		Non-Implementors n=108		
	Trained n=57	Not Trained n=26	Trained n=52	Not Trained n=56	
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F (p of F)
Innovative Instructional Methods	3.91 (1.26)	3.00 (1.11)	2.96 (1.27)	2.63 (.93)	12.52 (.000)
Estimated Hours	31 hrs.	16 hrs.	16 hrs.	11 hrs.	
Alternative Assessment Methods	2.98 (1.28)	1.00 (.00)	2.11 (.79)	1.00 (.00)	67.70 (.000)
Estimated Hours	16 hrs.	0 hrs.	8 hrs.	0 hrs.	

Believing that the level of encouragement by district administrator could influence teachers' perceived use of certain behaviors, teachers were asked to respond to a series of questions related to the extent that they believed their district encouraged the use of certain innovative instructional practices. Charting the responses by group (Figure 4) indicated patterns of perceived encouragement. Most of the differences were statistically significant at or above the .05 level. In general, alternative assessment trained teachers, Trained Implementors and Trained Non-implementors,

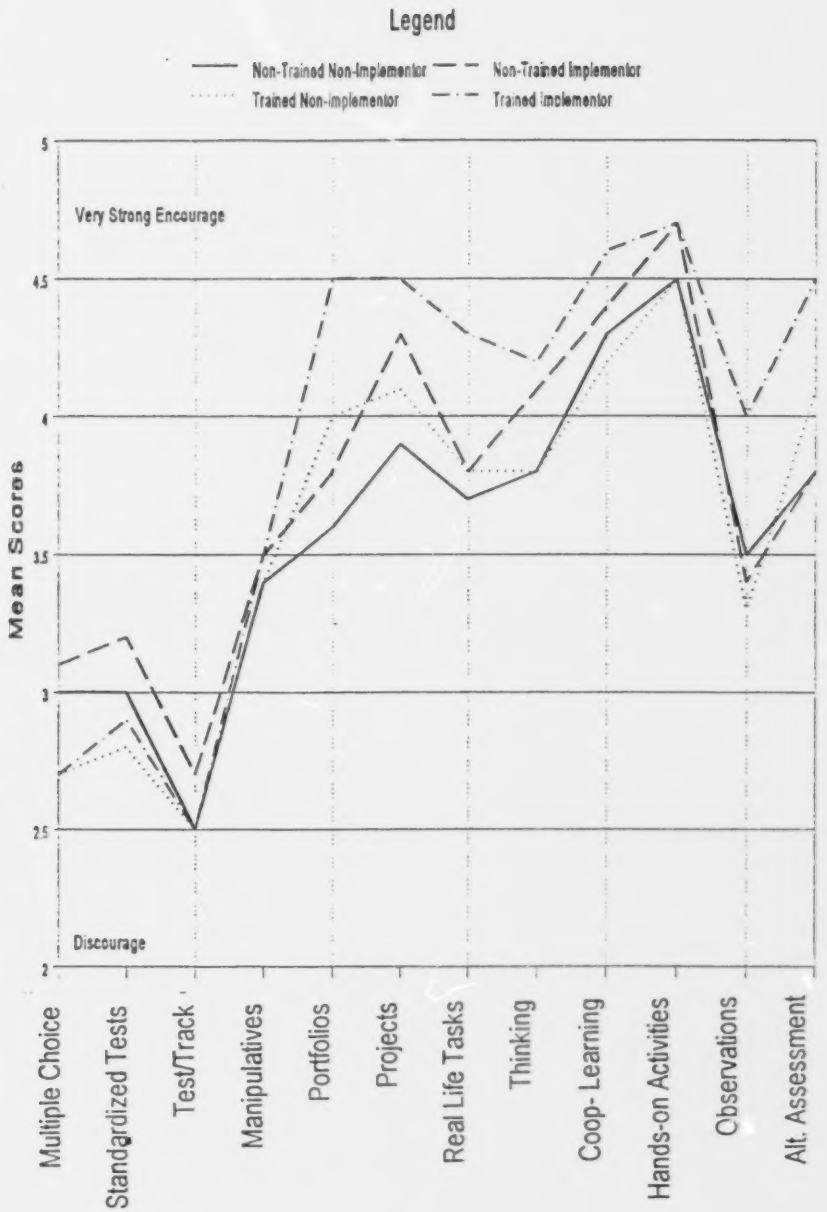

Figure 4. Mean Score Responses on Items Related to Level of Encouragement

reported lower levels of encouragement to use more traditional practices like multiple choice, standardized tests, and tracking. With the exception of encouragement to use portfolios and alternative assessment, Trained Non-implementors scored lower than most other groups on the level of encouragement to use most of the innovative assessment practices. *T*-test comparisons indicated that Trained Non-implementors' lower mean scores were statistically significant when compared to Trained Implementors on all of the items related to the encouragement to use innovative assessment practices, i.e. portfolios, group projects, real life tasks, explanations of thinking, cooperative learning, hands on activities, observations.

Beliefs Regarding Present Assessment System

Another factor that could influence responses reflecting implementation was the belief patterns that teachers reported regarding their present system of evaluation. It was my hunch that teachers who were dissatisfied with the system of assessment they were asked or encouraged to implement would exhibit lower levels of agreement with the beliefs posed regarding the ability of their system to function fairly. When asked to react to a series of statements regarding feelings about their present system of student assessment, Trained Non-implementors consistently responded lower than the other groups (Figure 5). The only item they responded more positively on was the item that asked if their system of evaluation "is driven by outside forces." This pattern could be interpreted to indicate that Trained Non-implementors: 1) were being influenced to implement a new system that they were not comfortable with; 2) had not

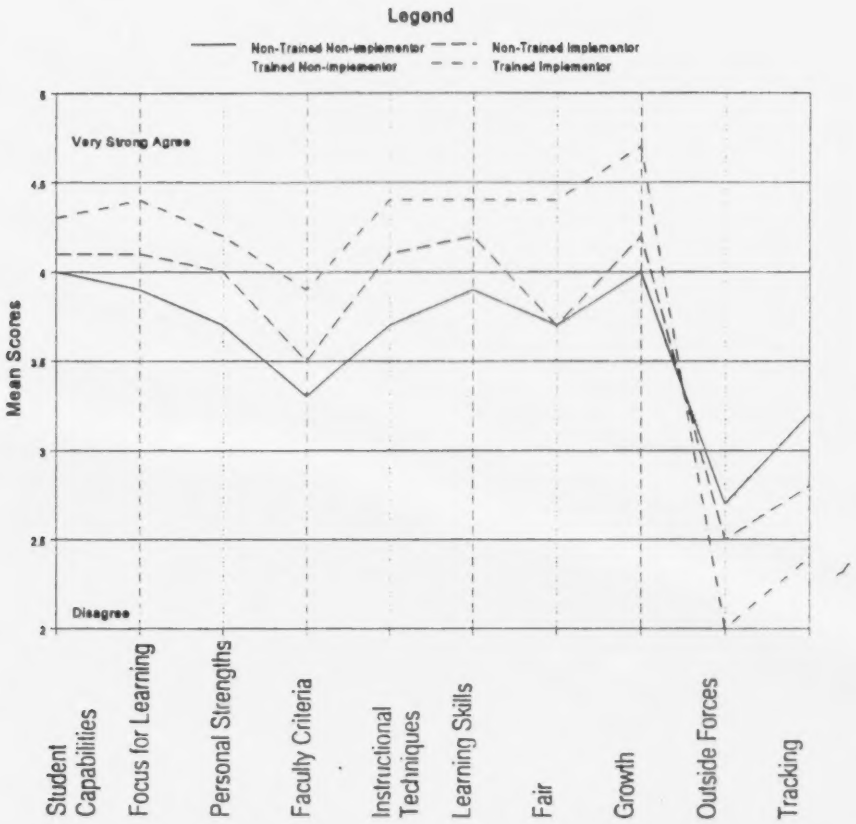

Figure 5. Mean Score Responses on Items Related to Beliefs About Present Assessment System

worked with the new system long enough to have positive feelings or beliefs about it; or 3) were uncomfortable with the old system but were experiencing resistance from administration or peers to change to the new system.

Differences in the Reported Use of Instructional Strategies

Depending on the amount of training and level of implementation, teacher responses reflected the components of alternative assessment (Table 5). Although

Table 5

Comparison of Means Scores on Component Indexes

(N=191)

	Implementors N=83				Non-Implementors N=108					
	Trained n=57		Not Trained n=26		Trained n=52		Not Trained n=56			
Component:	M	SD	M	SD	M	SD	M	SD	F	p of F
Student focused	2.9	.63	2.6	.80	2.0	.80	2.0	.60	32.51	.000
Longitudinal	3.6	.48	3.4	.57	2.8	.60	2.7	.74	25.70	.000
Integrates knowledge	4.1	.41	3.8	.48	3.3	.54	3.1	.64	36.30	.000
Non-Biased	3.6	.49	3.3	.45	2.9	.48	2.7	.60	31.64	.000
Process-oriented	3.4	.42	3.2	.60	2.7	.45	2.6	.56	27.96	.000
Authentic	3.6	.65	3.4	.56	3.1	.56	2.6	.82	22.01	.000
Not Secretive	3.1	.58	2.7	.66	2.3	.45	2.2	.65	26.84	.000

initial analyses indicated statistically significant differences between teachers trained in alternative modes of assessment and teachers without training in assessment, further analysis indicated that only 52% of the teachers with training were identified as implementing representative practices and strategies in their classroom. On the other hand, a small group (30%) of the teachers without assessment training, Non-trained Implementors were identified through their responses, as implementing related assessment practices.

In addition to component indexes items were analyzed according to topical groups as they related to traditional and assessment related behaviors. Mean score responses for items grouped according to topic are summarized in Table 6. The following dialogue summarizes the findings of specific items within both the component and topical grouping tables. All items discussed displayed statistically significant differences in ANOVA analyses ($p < .001$) unless otherwise stated.

The type of assessment practices reported by teachers reflected the focus of the evaluation. Teachers implementing alternative modes of assessment indicated through their responses a greater degree of concern with improvement, thinking process, application of knowledge and personal growth. As a result they reported using a greater variety of assessment techniques. When testing, they reported more frequent use of essays, open-ended questions, questions requiring explanations or justification of answers, complex problems and questions where students assessed their own learning. They were more likely to report the use of portfolios, projects, presentations and interdisciplinary instructional units that evolve over time and require the integration of

Table 6

ANOVA Comparison of Mean Summary Scores for Item Groupings

(N=191)

	Implementors N=83				Non-Implementors N=108					
	Trained n=57		Not trained n=26		Trained n=52		Not Trained n=56			
Item	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	F	p of F
Traditional Practices	2.89	.63	3.34	.48	2.95	.63	3.35	.56	8.01	.000
Test Items	3.90	.49	3.54	.55	3.25	.56	3.06	.63	22.90	.000
Task Requirements	3.72	.46	3.35	.47	2.96	.47	2.75	.65	36.10	.000
Grouping Behaviors	3.75	.51	3.68	.51	3.32	.53	3.30	.53	10.67	.000
Traditional Grouping Behaviors	2.31	.70	2.72	.67	2.56	.50	2.55	.61	4.78	.003
Presentation of Information	4.25	.48	4.07	.47	3.66	.58	3.60	.57	18.12	.000
Student Interaction	2.88	.59	2.49	.61	2.02	.39	2.02	.57	32.14	.000

knowledge. Interviews, observations and reflection activities were reported used to document classroom activity. These types of activities can give insight into the student's use of thinking skills or identify individual student needs.

As indicated in the self-reported behaviors, the implementation of alternative modes of assessment appears to have an influence on the relationship between teacher and student, changing the environment and organization of the classroom. Teachers who are implementing alternative modes of assessment more often reported the use of practices that give students the opportunity to make decisions regarding issues of instruction and assessment. For example, they reported that students were permitted to decide how to complete an assignment or to design their own research projects, grading criteria was discussed with the students both prior to and during the evaluation process. These types of reported behaviors encourage the revision and improvement of work and the internalization of expectations by the student. The format of the implementing classroom should reflect less teacher direction. Trained Implementors reported spending almost as much time in small groups with the teacher as facilitator or outside observer (23%) as they did in whole class instruction (26%). On the other end of the spectrum, the Non-trained Non-implementing teachers reported spending 41% of classroom time in whole class instruction. (Table 7).

Initial training in alternative modes of assessment appears to have some influence on the instructional strategies self-reported by teachers. Although Non-implementors did self report the use of more traditional strategies less than the teachers without training in alternative modes of assessment (Table 8), their responses

Table 7

Percentage of Time in Various Instructional Formats

(N=191)

	Implementors N=83		Non-Implementors N=108	
	Trained	Not trained	Trained	Not trained
	n=57	n=26	n=52	n=56
Item	M%	M%	M%	M%
Whole class instruction	26	29	34**	41**
Individual work by students	19	19	18	16
Teacher directed small group	13	15	14	15
Small groups / teacher intervention	16	16	15	12*
Small groups / teacher facilitator	23	21	18	14**

*p<.05, two-tailed **p<.01, two-tailed

did not indicate the move toward student involvement in the learning process (Table 9). They scored lower than Trained Implementors on items related to self assessment and self awareness, using and developing criteria with students or other teachers, and those practices related to integrated learning. Compared to Trained Non-implementors, Non-trained Implementors reported more frequent use of portfolios both in writing and in the content areas, documentation of thinking and student needs through interviews and more frequently requiring students to of self assess. Although they reported more frequent discussions of criteria with teachers and students, they did not report working more frequently with teachers to develop criteria. They also did not offer students the opportunity of demonstrating achievement through activities like exhibitions or

Table 8

ANOVA Comparison of Responses on Items Related to Traditional Assessment Practices

(N=191)

	Implementors N=83				Non-Implementors N=108					
	Trained n=57		Not trained n=26		Trained n=52		Not Trained n=56			
Item	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	F	p of F
Uses Texts, Workbooks, Skill Sheets	2.5	1.14	3.2	1.04	2.8	1.26	3.4	.93	7.43	.000
End of Unit Written Test	3.0	1.18	3.4	1.06	2.9	1.35	3.7	1.27	4.16	.007
Independent Tasks	3.5	.83	3.4	.64	3.3	.61	3.7	.72	2.81	.041
Work/Ability Group	2.4	1.30	3.0	1.06	2.4	1.20	2.9	1.20	2.72	.046
Ability Group/Track	2.1	1.11	2.6	1.06	2.4	1.04	2.6	1.07	2.27	.082
Rank on Curve	1.8	1.05	2.3	.89	1.6	1.35	1.9	1.20	2.80	.041
Short Answers	2.8	1.01	3.1	.81	3.1	.96	3.3	.80	2.76	.004
Direct Instruction	3.4	.74	3.8	.82	3.5	.64	3.8	.71	5.18	.000
Secures Questions	2.5	1.38	3.6	1.32	2.5	1.46	2.7	1.54	4.30	.006

Table 9

Comparison of Responses for Implementors and Non-Implementors on Items Related to Student Interaction

(N=191)

	Implementors N=83				Non-Implementors N=108					
	Trained n=57		Not Trained n=26		Trained n=52		Not Trained n=56			
Item	M	SD	M	SD	M	SD	M	SD	F	p of F
Critique peers	2.5	0.93	2.1	1.02	1.5	0.7	1.7	0.97	11.47	.000
Tutor others	2.8	1.09	2.8	1.13	2.2	1.04	2.1	1.04	4.93	.003
Record thoughts	2.7	1.15	2.3	1.18	1.6	0.75	1.7	1.05	13.19	.000
Self assess	2.7	0.97	2.3	1.01	1.9	0.92	1.8	1.04	10.62	.000
Self awareness	3.1	0.97	2.7	0.87	2.3	0.86	2.2	0.82	10.48	.000
Set goals	3.4	0.7	3.2	0.9	2.5	0.97	2.7	0.98	10.97	.000
Compose questions	3.5	1.17	3.2	1.13	2.3	1.06	2.4	1.1	15.58	.000
Design research	2.6	1.01	2.2	1.1	1.7	0.93	1.6	0.64	13.55	.000
Decide how	3.4	0.78	3	0.87	2.7	0.85	2.5	1.03	10.74	.000

presentations, or to demonstrate application by having students compose possible questions as well as answer them.

Factors Influencing Implementation

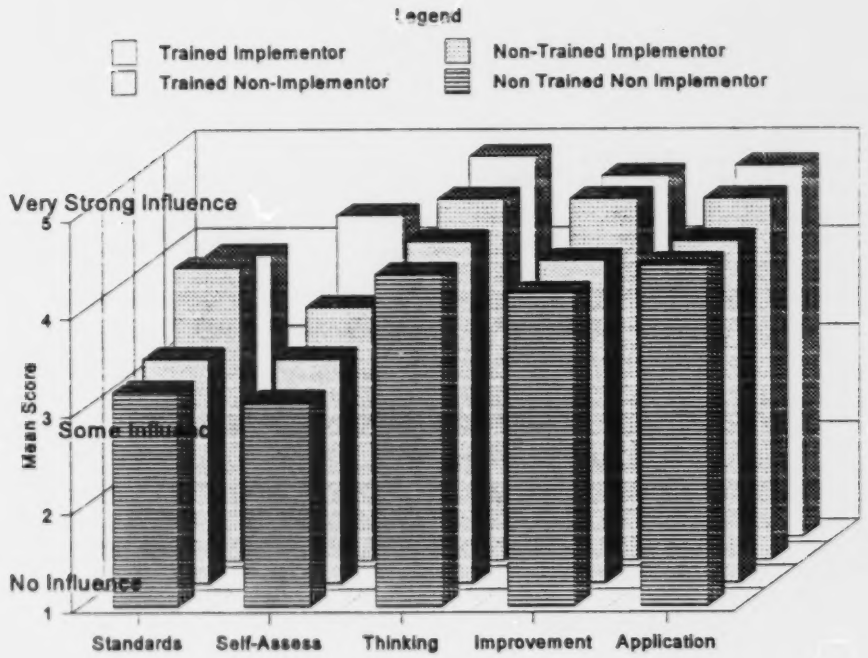

Figure 6 Reported Influences of Assessment Evaluation and Design

Although the amount of training in alternative modes of assessment and innovative instructional methods could have been a factor differentiating Trained Implementors and Non-implementors, it was difficult to ascertain what demographic or extraneous factors were related to Non-trained Implementors responses. Non-

trained Implementors and Trained Non-implementors reported similar training in innovative instructional methods. *T*-test comparisons between Non-trained Implementors and Trained Non-implementors, resulted in statistically significant differences in mean score comparisons on items related to the influence of district standards of performance for every child and the influences on grading from the final product and improvement (Figure 6). The Trained Non-implementors' reported lower level of influence by district standards and guidelines could restrict the use of strategies that have students self assess, tutor or discuss criteria, while high influence in the same area could promote their use. In addition, teachers identified as implementing also reported more frequently developing and discussing criteria with

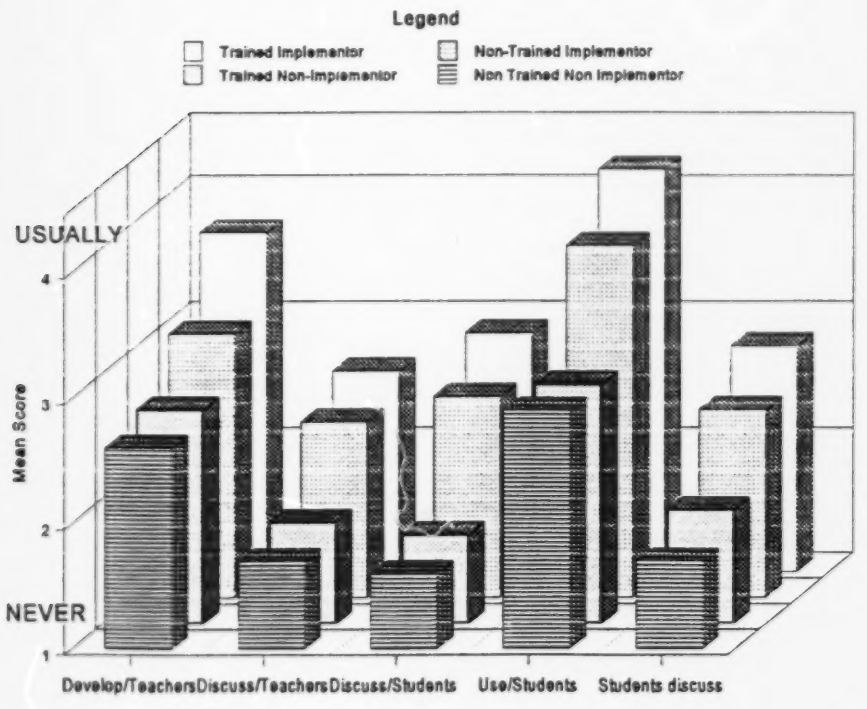

Figure 7. Development and Use of Criteria for Assessment and Evaluation

other teachers as well as encouraging students to discuss and use the established criteria (Figure 7). This reported behavior appears to reflect the presence of common expectations, clear standards and collaborative environment. Therefore, the level of training combined with levels of perceived encouragement, beliefs about their present system of assessment, levels of influence by district standards and collaborative work on the development and use of criteria could have affected the teachers' ability to implement strategies related to alternative modes of assessment. This premise finds support in the literature previously discussed on staff development which contends that implementation is supported by a shared mission, collaborative goals, and a supportive environment (Loucks-Horsley & Stiegelbauer, 1991; McLaughlin, 1991).

Limitations of the Study

There were three limitations to be considered when reviewing this study. One, the findings were based on self-reports by teachers regarding what they perceived to be their behaviors. Two, the results are most representative of elementary teachers (grades 1-6) in the state of New York. And, third, findings were based on reports from teachers working primarily in suburban school districts with per pupil expenditures that exceed the average for suburban districts in the designated state.

As with any survey, the results of this study were based on the perception of teachers as they responded to the questions. Whether the teachers were actually implementing the behaviors as described or were repeating the rhetoric of the methodology can only be determined through actual classroom observations over

extended periods of time. Nevertheless, the consistency of responses over similar items, (i.e. use of portfolios, use of interviews) asked in a variety of formats increased the possibility that teacher responses were reflective of behavior patterns rather than rhetoric. Also, for the purposes of this study, the responses, even if they represent the rhetoric of the discipline, reflect those aspects teachers appear to be internalizing. It is important to know which aspects of training experiences teachers remember or report back and which they do not.

Significance of the Study

Implications for Research

This study expanded on the findings of Aschbacher (1993), Koretz et al. (1992) and Vitali (1993) regarding the perceived effects of performance assessments and portfolios on instruction and/or teacher attitudes. In addition, this study takes a beginning step toward bridging the gap between the subtle differences in training content and technique by identifying commonalities and constructs. It also takes an initial step toward the exploration of the influences of training, school environment and school culture on implementation.

Additional research is needed to clarify many of the issues and to answer new questions that have been raised in these findings. A series of on-site visitations, interviews and observations could provide additional qualitative data to support the findings of this study regarding the perception of student involvement in the process of learning, the establishment of criteria and the types of assessments used. A closer

qualitative analysis of teachers' behaviors would also give insight into issues of bias and equity as they relate to issues of culture, multiple intelligence and learning styles.

Implications for School Districts

Proponents of alternative modes of assessment argue that just as testing for lower order thinking skills has influenced teachers' instructional strategies, a better test will influence the use of better teaching strategies (Archbald, 1991; Wiggins, 1990; Wolf et al.,1991). The self-reported testing practices and strategies of Trained Implementors' reflect an emphasis on the thinking skills, complex problems, student performance over time, and the application rather than recall of knowledge that researchers have related to classroom reform (Darling-Hammond, 1990,1993; Simmons & Resnick, 1993; Wiggins, 1991). If classroom reform is to be an option of the new assessment methodologies school districts need to provide adequate and on-going training, administrative support, and to establish collaborative goals. Aschbacher (1992), McLaughlin (1991), Lieberman & Miller (1991) have indicated that one of the key barriers to implementation is lack of time to learn, plan, practice, use and reflect. Repeatedly advocates of alternative modes of assessment have warned that it is a time intensive process. The observation that most implementors reported in excess of 30 hours of training in alternative assessment and in the identified innovative instructional practices, the fact that they indicated frequent discussions with other teachers, reported strong influence from district standards of performance and that perceived high levels of encouragement from district administrators may indicate that these

teachers had access to the amount of training, type of support, assistance and shared vision needed to foster desired changes.

Conclusion

Significant differences were found between the instructional strategies and practices of teachers and implementing and those not implementing alternative modes of assessment. An analysis of those strategies and practices reflects the theories advocated by supporters of assessment reform. Implementing teachers self-reported strategies seem to incorporate the use of research-based behaviors and practices believed to positively effect student learning, application of knowledge and self-regulated behavior into their classroom environment. Assessment appears congruent with instruction. Nevertheless, without proper training, perceived support for training, encouragement and a shared-mission these practices may not become part of the teachers' reported repertoire of practice.

BIBLIOGRAPHY

Archbald, D. A. (1991). Authentic assessment: What it means and how it can help schools. Madison,WI: University of Wisconsin-Madison, National Center for Effective Schools Research and Development.

Aschbacher, P., (1991). Alternative assessment: State activity, interest, and concerns (CSE Tech. Rep. No. 322). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

_____, (1993). Issues in innovative assessment for classroom practice: Barriers and facilitators (CSE Tech. Rep. No. 359). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E. L., & Stites, R. (1991). Trends in testing in the USA. In S. Fuhrman & B. Malen (Eds.). The Politics of Curriculum and Testing (pp. 139-158). New York: Falmer.

Baron, J. B. (1990, July). Toward a new generation of student outcome measures: Connecticut's common core of learning assessment. Unpublished manuscript, Connecticut State Department of Education, Division of Research, Evaluation and Assessment. Paper presented at New Modes of Assessment Conference, Harvard Graduate School of Education, Cambridge, MA.

Caine, R., & Caine, G. (1992) Making connections: Teaching and the human brain. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (1990, December). Achieving our goals: Superficial or structural reforms? Phi Delta Kappan, pp. 286-295.

_____,. (1992). Interim Report of the Committee on Curriculum and Assessment. New York State Education Department. Albany, NY: University of the State of New York.

_____,. (1993, June). Discussion at New York State Educational Leadership Conference, Albany, N.Y.

Frederiksen, J. R., & Collins, A. (1989, December). A systems approach to educational testing. Educational Researcher, pp. 27-32.

Fullan, M. & Miles, M., (1992, June). Getting reform rignt: What works and what doesn't. Phi Beta Kappan, 73, pp. 745-752.

Gardner, H. (1985). Frames of Mind: The Theory of Multiple Intelligences. New York: Basic Books, Inc.

_____, H. (1989). Assessment in context: The alternative to standardized testing. In Bernard R. Gifford (Ed.) Report of the Commission on Testing and Public Policy. Boston: Kluwer Academic Publishers.

Gardner, H., & Hatch, T. (1989, November). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. Educational Researcher, pp. 4-10.

Gooding, K., (1992, January). Alternative Modes of Assessment. Unpublished manuscript. Paper presented Alternative Assessment: Using Standards, Not Standardization, workshop presented by Nassau Board of Cooperative Educational Services, Department of Research Evaluation and Planning, Westbury, NY.

Hall, G. & Loucks, S. (1979) Teacher concerns as a basis for facilitating and personalizing staff development. In Lieberman, A. & Miller, L. (Eds.). (1979). Staff Development: New Demands, New Realities, New Perspectives, (pp. 36-53) New York: Teachers College Press.

Herman J., Aschbacher, P., & Winters, L. (1992). A practical guide to alternative assessment. Alexandria, VA: Association for Supervision and Curriculum Development

Herman, J., Dreyfus, J., & Golan, S. (1990). The effects of testing on teaching and learning (CSE Tech. Rep. No. 327). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

_____, & Golan, S. (1991). Effects of standardized testing on teachers and learning Another look (CSE Tech. Rep. No. 334). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

_____. (1992, February) What's happening with educaional assessment? Tallahassee, FL: South Eastern Regional Vision for Education (SERVE)

Koretz, D., Stecher, B., & Deibert, E. (1992, August). The Vermont Portfolio Assessment program: Interim report on implementaiion and impact,1991-92 school year (CSE Tech. Rep. No. 350). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Neill, M. & Medina, N. (1989, May). Standardized testing: Harmful to educational health. Phi Delta Kappan, pp. 88 - 697.

New York State Education Department. (1991). A New Compact for Learning: Improving Public Elementary Middle, and Secondary Education Results in the 1990s. Albany, NY: University of the State of New York.

New York State Education Department. (1992). New York State learning: Statistical Profiles of public school districts. Albany, NY: University of the State of New York.

New York State United Teachers (NYSUT) (1991). Multiple choices: Reforming student testing in New York State. Albany, NY: New York State United Teachers/American Federation of Teachers, AFL-CIO.

Perrone, V. (Ed.). (1991). Expanding Student Assessment. Alexandria, VA: Association for Curriculum and Development.

Resnick, L. B. & Resnick, D. P. (1989a). Assessing the thinking curriculum: New tools for educational reform. To appear in B. R. Gifford, & M. C. O'Connor (Eds.), Future assessments: Changing views of aptitude, achievement, and instruction. Boston: Kluwer Academic Publishers.

_____ & _____. (1989b). Tests as standards of achievement in schools. The uses of standardized tests in American education: Proceedings of the 1989 ETS Invitational Conference (pp. 63-80). Princeton, NJ: Educational Testing Service.

Resnick, L. B., & Klopfer, L. E. (Eds.). (1989). Toward the thinking curriculum: Current cognitive research. Alexandria, VA: Association for Supervision and Curriculum Development.

Rosenholtz. S (1989). Teachers' Workplace: The social organization of schools. New York: Longman.

Shepard, L. A. (1991). Readiness testing in local school districts: An analysis of backdoor policies. In S. Fuhrman & B. Malen (Eds.), The Politics of Curriculum and Testing (pp. 159-180). New York: Falmer.

Simmons, W., & Resnick, L., (1993, February). Assessment as a catalyst of school reform. Educational Leadership, pp. 11-15.

Sternberg, R. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

LeMahieu, P. & Leinhardt, G. (1989). Overlap: Influencing what's taught. A process model of teachers' content selection. Journal of Classroom Interaction, 21(1), 2-11.

Lieberman, A. & Miller, L. (Eds.). (1991). Staff Development for Education in the '90s: New Demands, New Realities, New Perspectives. New York: Teachers College Press

_____. (1991). Revisiting the Social Realities of Teaching. In Lieberman, A. & Miller, L. (Eds.). (1991). Staff Development for Education in the '90s: New Demands, New Realities, New Perspectives, (pp. 92-109) New York: Teachers College Press

LeMahieu, P. G., & Wallace, R. C., Jr. (1986). Up against the wall: Psychometrics meets praxis. Educational Measurement: Issues and Practice, 2, 12-16.

Linn, R. E., (1992). How we know when assessment works. Presentation at What Works in Performance Assessment, National Center for Research on Evaluation Standards and Student Testing (CRESST) Conference, Los Angeles.

_____, Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

Loucks-Horsley, S. & Stiegelbauer, S. (1991). Using Knowledge of change to guide staff development. In Lieberman, A. & Miller, L. (Eds.). (1991). Staff Development for Education in the '90s: New Demands, New Realities, New Perspectives, (pp. 92-109) New York: Teachers College Press.

Lott, W.A. (1991, May). Student assessment in the 1990s. New York School Boards Journal, pp. 36-42.

McLaughlin, M.W. (1991). Enabling Professiional Development: What we have learned? In In Lieberman, A. & Miller, L. (Eds.). (1991). Staff Development for Education in the '90s: New Demands, New Realities, New Perspectives, (pp. 61-81) New York: Teachers College Press

Mitchell, R. (1992). Testing for learning: How new approaches to evaluation can improve American schools. New York: The Free Press.

National Commission on Testing and Public Policy. (1990). From gatekeeper to gateway: Transforming testing in America. Massachusetts: Boston College, National Commission on Testing and Public Policy.

Stiggins, R. (1987, Fall). Design and development of performance assessments. Educational Measurement: Issues and Practice, pp. 33-42.

_____. (1989, May). Assessment Literacy. (Available from Northwest Regional Educational Laboratory, 101 S. W. Main, Suite 500, Portland, Oregon 97204).

Vitali, G. (1993). Factors influencing teachers' assessment and instructional practices in an assessment-driven educational reform. Umpublished doctoral dissertation, University of Kentucky, Kentucky.

Wiggins, G. (1989, May). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, pp. 703-713.

Wiggins, G. (1990). The case for authentic assessment. (Report No. EDO-TM-90-10). Washington, D.C.: American Institutes for Research, Eric Clearinghouse on Tests, Measurement, and Evaluation. (ERIC Document Reproduction Service No. ED 328 611).

_____. (1991, October 15). Teaching to the (authentic) test: Assessment that measures and evokes quality student performance. Presentation at Harvesting Connections Conference, New York State, Association for Supervision and Curriculum Development and the New York State Education Department, Ellenville, NY.

Wolf, D. P. (1987,December/1988, January). Opening up assessment Educational Leadership, pp. 24-29.

_____. (1989, April). Portfolio assessment: Sampling Student Work. Educational Leadership, pp. 35-39.

_____., Bixby, J., Glenn, J., III, & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. Review of Research in Education, 17, 31-74.

_____, LeMahieu, P., & Eresh, J. (1992, May). Good measure: Assessment as a tool for educational reform. Educational Leadership, pp. 8-13.